DOCUMENT RESUME

ED 469 186 TM 034 491

AUTHOR	Bolt, Daniel; Roussos, Louis; Stout, William
TITLE	Estimation of Item Dimensional Measurement Direction Using Conditional Covariance Patterns. Computerized Testing Report. LSAC Research Report Series.
INSTITUTION	Law School Admission Council, Newtown, PA.
REPORT NO	LSAC-R-98-02
PUB DATE	2002-07-00
NOTE	27p.
PUB TYPE	Numerical/Quantitative Data (110) -- Reports - Research (143)
EDRS PRICE	EDRS Price MF01/PC02 Plus Postage.
DESCRIPTORS	Analysis of Covariance; Estimation (Mathematics); Item Response Theory; Scaling; *Test Items
IDENTIFIERS	Conditionals; *Dimensionality (Tests)

ABSTRACT

Several nonparametric dimensionality assessment tools have demonstrated the usefulness of item pair conditional covariances as building blocks for investigating multidimensional test structure. Recently, J. Zhang and W. Stout (1999) have related the structural properties of conditional covariances in a generalized compensatory framework to a test item geometry model analogous to that developed by M. Reckase (1985) for the multidimensional logistic item response function. This model represents an item's measurement direction in terms of the multidimensional ability composite best measured (in terms of information) by the item. In this preliminary study, it is demonstrated how such directional representations can be reconstructed on the basis of each item's individual pattern of conditional covariances with all of the other items on the test. A proximity measure for item pairs based on these patterns is derived, and circular and spherical multidimensional scaling techniques are shown to provide a method by which the unique measurement directions of items might be recovered and represented. Several real and simulated data analyses suggest that the approach may be a useful too, for investigating and describing the dimensional structure of tests. (Contains 4 tables, 12 figures, and 24 references.) (Author/SLD)

ED 469 186

LSAC RESEARCH REPORT SERIES

TM

■ Estimation of Item Dimensional Measurement Direction Using Conditional Covariance Patterns

Daniel Bolt
University of Illinois at Urbana-Champaign

Louis Roussos
Law School Admission Council

William Stout
University of Illinois at Urbana-Champaign

■ Law School Admission Council
Computerized Testing Report 98-02
July 2002

TM034491

LSAC®

A Publication of the Law School Admission Council

The Law School Admission Council is a nonprofit corporation that provides services to the legal education community. Its members are 200 law schools in the United States and Canada.

This study is published and distributed by the Law School Admission Council (LSAC). The opinions and conclusions contained in these reports are those of the authors and do not necessarily reflect the position or policy of the Law School Admission Council.

Table of Contents

Executive Summary

As proposed in the Computerized Law School Admission Test (LSAT) Research Agenda, the Law School Admission Council (LSAC) is currently investigating the advisability and feasibility of developing a computerized version of the LSAT. An important component of this research effort is the development of new dimensionality estimation procedures. For the purposes of this report, the dimensions of a test may be thought of as the number of statistically detectable skills that the test is measuring. A dimensionality analysis involves determining the number of dimensions (or skills) being measured by the test, the nature of these dimensions, and the degree to which the dimensions are correlated.

The current LSAT has been reliably estimated by several studies to have two dominant dimensions. One of these dimensions is associated with the analytical reasoning (AR) items, while the other is associated with the combination of logical reasoning (LR) and reading comprehension (RC) items. However, previous analyses have also demonstrated other more minor dimensions, for example, the item sets corresponding to the different AR and RC passages. Moreover, in a computerized adaptive test (CAT) setting, new minor dimensions may be introduced because of the new medium of administering the test or because of the adaptive way in which the items are administered. The influence of such minor dimensions are strictly controlled on the current LSAT so that no test takers are unfairly advantaged or disadvantaged. To ensure the continued control over these minor dimensions, new dimensionality estimation techniques should be developed for CAT data so that we can continue to monitor the influence of such dimensions in a CAT setting.

The purpose of the current paper is to develop a new more powerful dimensionality estimation technique. The new technique expands upon previously developed techniques that could only determine whether two items were measuring the same dimension by also measuring the degree of dimensional similarity between two items.

Through analysis of simulated data, the new technique is shown to be able to estimate dimensional similarity between items that previous techniques were unable to measure. Perhaps the most striking results with the new technique were demonstrated in an analysis of LSAT RC data. In this analysis the new technique not only resolved the conflicting results from two previous analyses, it also yielded a clear and compelling explanation as to why the conflict had occurred.

In conclusion, the new technique introduced in this paper shows much promise as a potentially powerful dimensionality estimation tool.

Abstract

Several nonparametric dimensionality assessment tools have demonstrated the usefulness of item pair conditional covariances as building blocks for investigating multidimensional test structure. Recently, Zhang and Stout (1999) have related the structural properties of conditional covariances in a generalized compensatory framework to a test item geometry model analogous to that developed by Reckase (1985) for the multidimensional logistic item response function. This model represents an item's measurement direction in terms of the multidimensional ability composite best measured (in terms of information) by the item. In this preliminary study we demonstrate how such directional representations can be reconstructed on the basis of each item's individual pattern of conditional covariances with all of the other items on the test. A proximity measure for item pairs based on these patterns is derived, and circular and spherical multidimensional scaling techniques are shown to provide a method by which the unique measurement directions of items might be recovered and represented. Several real and simulated data analyses suggest that the approach may be a useful tool for investigating and describing the dimensional structure of tests.

Introduction

As most educational and psychological tests are recognized as being at least to some degree multidimensional, the development of accurate dimensionality assessment tools has become increasingly important. These tools serve the purposes both of identifying when test multidimensionality exists and of providing a practical, useful description of multidimensional test structure when it is known (or can be shown) to exist. This latter purpose can have important implications for several aspects of testing. Perhaps the most immediate is its ability to suggest an interpretation of the dimensions contributing to test performance. This is especially critical in cases where the practitioner may be uncertain as to whether such dimensions are intended to be measured by the test. For conditions where the test is intended to be multidimensional, an understanding of its multidimensional structure may also be important to the test item writer who can be made sensitive to how various types of items measure particular dimensions. For the psychometrician interested in item response theory (IRT) applications, test data analysis often depends on a unidimensionality assumption that, if not satisfied for the test as a whole, may nevertheless be satisfied by particular subsets of test items. The ability to recognize subgroups of items within a test as dimensionally homogeneous may permit use of IRT for those item subsets.

Most important from the perspective of the Law School Admission Council (LSAC), dimensionality assessment also becomes crucial in the context of computerized adaptive testing (CAT) when the pool of test items is known to measure more than one dimension. Because in CAT each examinee is administered a limited number of potentially unique test items, it is crucial from a test fairness perspective that the administered items adequately measure each of the intended abilities across examinees to the same extent. Thus, item selection algorithms may need to be sensitive in part to the degree that items measure each ability dimension. Because LSAC is in the midst of a five-year program to assess the feasibility and advisability of developing a computerized version of the Law School Admission Test (LSAT) (Pashley, 1995,1997), the development of improved dimensionality estimation procedures at LSAC is currently of particularly special interest.

Among the dimensionality assessment methods that have been previously developed is a class of methods that are characterized as *parametric*. Parametric methods explore multidimensionality by assuming a specific model for the item response function, a function that in the context of multidimensionality may be based on more than one ability dimension. Examples of parametric techniques include the application of linear factor analysis to a tetrachoric correlation matrix (Knol & Berger, 1991; Lord, 1980) and nonlinear factor analysis (McDonald, 1967). With these methods, an interpretation of multidimensional test structure is derived from the estimated factor loadings of the model. That is, items having a similar pattern in their loadings on the factors can be thought of as dimensionally similar.

One natural limitation to parametric methods is their strong modeling assumptions. In an attempt to relax these assumptions, several more recently developed non-parametric assessment tools emphasize the use of item pair conditional covariances (given a test composite score) as building blocks for investigating multidimensional test structure (see Stout, Habing, Douglas, Kim, Roussos & Zhang, 1996). The relevance of conditional covariances to dimensionality assessment is made especially apparent by recognizing the close association between local independence (LI) and dimensionality (Stout, 1990), and that in virtually all practical cases conditional covariances are sufficient for assessing LI (McDonald, 1985). One realization of this is Stout's DIMTEST procedure (Stout, 1987) for testing essential unidimensionality, which ultimately may be seen as testing for positive conditional covariances within an assessment subtest of items given an appropriately chosen partitioning subtest. Another example is Roussos' (1992) HCA/CCPROX, which sorts and identifies dimensionally similar item subsets by performing a hierarchical cluster analysis using a conditional-covariance-based proximity measure between items. Similarly, the procedure DETECT (Kim, 1994; Zhang & Stout, 1999) optimizes an index defined using conditional covariances to determine both the "amount" of multidimensionality present in the test and the number of dominant dimensions, as well as the items that measure those dimensions. Both HCA/CCPROX and DETECT might be thought of as methods for searching for patterns in the item pair conditional covariance matrix indicative of some degree of item level simple structure.

Recently Zhang and Stout (1999) have related the theoretical structural properties of conditional covariances given a latent test composite to a test item geometry model analogous to that developed by Reckase (see Reckase & McKinley, 1991). Adopting such a representation can in itself be a useful method for investigating and describing multidimensional test structure, as has been demonstrated with Reckase's multidimensional logistic model (Reckase, 1985, 1997). It also provides a valuable framework in which an interpretation of other dimensionality analysis tools, such as DIMTEST, HCA/CCPROX, and DETECT, may be enhanced. Zhang and Stout's (1999) theoretical results importantly suggest that well-estimated conditional covariances may in themselves, and apart from adoption of a particular multidimensional item response model, provide a basis for such a geometric description. In this paper, we more specifically investigate the utility of an item's pattern of conditional covariances with other items as an indicator of its underlying geometric structure. We begin by introducing a measure of dissimilarity between two items based on the correspondence of their conditional covariance patterns. Then, using this measure, we propose a class of multidimensional scaling techniques to construct an item-level geometric representation of their dimensional structure. We illustrate the potential of this method using several simulated data sets and then investigate the dimensional structure of two sections of the LSAT, the analytical reasoning (AR) and reading comprehension (RC) sections.

Representation of a Generalized Compensatory Model

For many tests that are multidimensional, it may be reasonable to assume that the skills or abilities being measured interact in a compensatory way. Qualitatively, a multidimensional compensatory model refers to one in which a relative weakness in one ability can be compensated for by relative strength in another ability. For example, on a math test an examinee lacking in geometry ability may nevertheless be able to answer a question measuring both geometry and algebra skills correctly because of high algebra ability. In this case high algebra ability would be considered to compensate for insufficient geometry ability. Multidimensional compensatory models may be especially relevant for tests where different strategies can be brought to bear in solving a test item. This may occur, for example, in a math test where an individual item might be solved by following either a geometric or algebraic strategy.

Zhang and Stout (1999) consider the case of a generalized compensatory model (GCM) in which all item response functions have the general form:

$$P_i(\theta) = H_i\left(a_i^T \theta\right), \quad i = 1, 2, \ldots, n$$

where $a_i = (a_{i1}, a_{i2}, \ldots, a_{in})$ is a discrimination parameter vector, and $H_i(x)$ is a nondecreasing function with

$$H_i^{'}(x) \geq 0 \text{ and } H_i^{'}(x) H_i^{'}(y) \text{ not being 0 identically, for } i, j = 1, 2, \ldots, n.$$

Specific parametric multidimensional models may be defined by varying the nature of the item response function in the above representation. Two common examples are the multidimensional logistic model of Reckase (1985) and the multidimensional normal ogive model discussed by McDonald (1985), where the H_is are based on the logistic and cumulative normal distribution functions, respectively. In the case of no guessing, Reckase's model can be formalized as follows:

$$P_i(\theta) = \frac{exp\left(\sum_{k=1}^{D} a_{ik}\theta_k + d_i\right)}{1 + exp\left(\sum_{k=1}^{D} a_{ik}\theta_k + d_i\right)}, \quad i = 1, 2, \ldots, n$$

where d is a multidimensional difficulty parameter such that larger values of d correspond to easier items, and D is the number of ability dimensions.

An example of an item vector representation that can be used in conjunction with a GCM is shown in Figure 1(a) for a set of 4 two-dimensional items. In keeping with the generalized nature of the GCM, the representation here is constructed as if independent of the functional form of H_i and depends only on the a_i for each individual item. In this case the a_i parameters are represented both by the direction of the vector and its length. Specifically, the angle between each item vector and the θ_1-axis corresponds to arctan

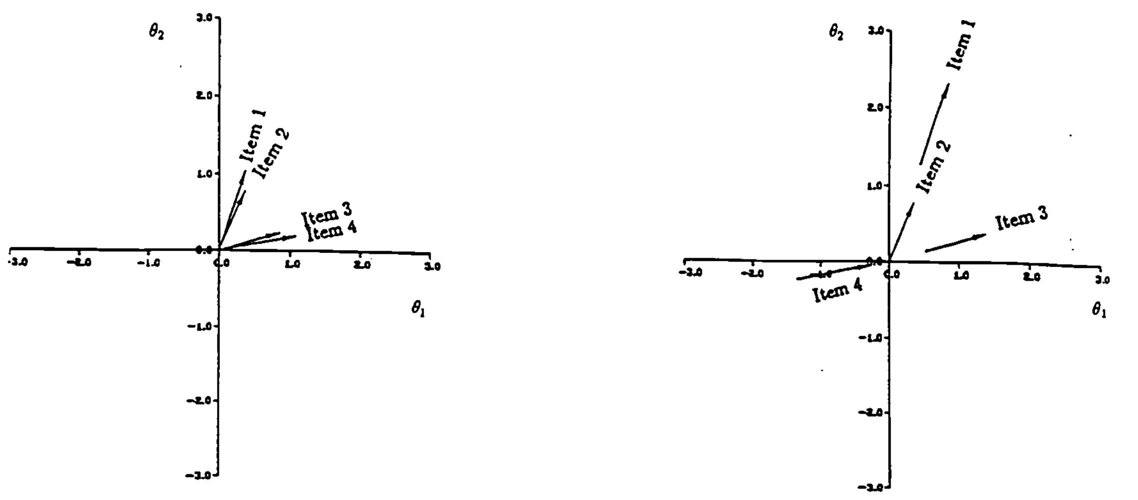
$$\frac{a_{i2}}{a_{i1}}$$

while the length of the vector is proportional to

$$\sqrt{\left(a_1^2 + a_2^2\right)}.$$

It will be useful to think of the direction of this vector as also representing the *measurement direction* of the item, as it corresponds to the unique direction in the latent space that maximizes multidimensional information (Zhang & Stout, 1999). Thus, items that measure a similar composite of dimensions will have similar measurement directions, and their vectors will lie in close angular proximity even though they may differ in terms of their lengths, or discriminating power. For the example in Figure 1(a), Items 1 and 2, which are oriented in very similar directions, also measure a similar dimensional composite, while Items 1 and 3 point in different directions and, thus, measure more distinct dimensional composites. Also, Item 1 is more discriminating than Item 2 with respect to their common measurement direction as it has a greater length.

Reckase (1985) added a representation of item difficulty based on the multidimensional item difficulty parameter d incorporated into the multidimensional logistic model. In this context, the tail of item i's vector is positioned d_i units from the origin in the direction of the item vector. Reckase (1985) shows that the position of the tail also corresponds to the location of the .5 probability contour of the item. Consequently, items that are more difficult will have tails positioned in the first quadrant while items that are easier will have tails located in the third quadrant. In Figure 1(b) the difficulty representation has been added, so that the order of difficulty of the items in terms of their multidimensional item difficulty is Item 1, followed by Items 3, 2, and 4, with Item 4 being the most difficult.



(a.) Representation of item discrimination direction and magnitude.

(b.) Representation of both difficulty level magnitude as well as discrimination direction and magnitude.

FIGURE 1. *Examples of item vector representations*

Within this representation, it is also useful to refer to a particular dimensional composite measured by a collection of multidimensional items, or an entire test. Specifically, the latent ability composite associated with a particular test (or subtest) composite score Y will be denoted as Θ_Y. The measurement direction in θ-space for this composite is defined as the direction for which Y provides the maximum expected information (Zhang & Stout, 1999). The direction in the latent space in which Θ_Y is measured is determined by the weights Θ_Y attached to the different dimensions. For example, in the case of two dimensions, θ_1 and θ_2, an equally weighted composite ($\Theta_Y = (\theta_1 + \theta_2)/2$) results in its measurement direction being exactly halfway between the θ_1 and θ_2 axes (i.e., at 45°).

Roughly speaking, the direction in which Θ_Y is measured might be thought of as some sort of an average direction based on the measurement directions of the items that comprise Y and in this sense is conceptually similar to Wang's (1988) notion of a reference composite. For details on defining multidimensional item or test information, or how the former combine to define the latter, see Zhang and Stout (1999). Most relevant for the purpose of this paper is the fundamental relation between the directions and lengths of the item vectors and their corresponding item-pair conditional covariances given Θ_Y, the focus of the next section. In our attempt to represent test structure through multidimensional scaling, it is ultimately these generalized compensatory model-based directions that we are attempting to recover.

Conditional Covariances and the GCM

As described above, associated with each item is a direction in the θ-space defined by its a_i vector, and associated with each test is an ability composite denoted by Θ_Y. We now consider the general case of a composite Θ_α, where $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_D)^t$, D is the number of dimensions, and the α_is denote the angles of the composite with each of the dimensions. In general, the α_is define how the different θ_i axes are weighted in the Θ_α composite with smaller values indicating greater weights. The following theorem provides a link between the conditional-covariance behavior of a pair of items and the angular difference in their measurement directions relative to a prescribed composite Θ_α.

Theorem (Zhang, 1996; Zhang & Stout, 1999): *If* $\Theta \sim N(0, I_D)$, $D \geq 2$, *and* $\Theta_\alpha = \alpha^T\Theta$ *is any fixed composite, then*

$$Cov\left(a_i^T\Theta, a_j^T\Theta \mid \Theta\alpha\right) = \left\|a_i^\perp\right\|\left\|a_j^\perp\right\| \cos\beta_{ij}, \tag{1}$$

$$= \left\|a_i\right\|\left\|a_j\right\| \sin\beta_i \sin\beta_j \cos\beta_{ij}, \tag{2}$$

where $a_i^\perp$ *and* $a_j^\perp$ *are the projections of the discrimination parameter vectors* a_i *and* a_j *on the space*

$$V_\alpha^\perp \overset{def}{=} \left\{y = \left(y_1, y_2, \ldots, y_D\right)^T : \alpha^T y = 0\right\}, \text{ respectively;}$$

β_{ij} *is the angle between* $a_i^\perp$ *and* $a_j^\perp$ $(0 \leq \beta_{ij} \leq \pi)$, $1 \leq i \neq j \leq n$; *and* β_i, β_j *are the angles between* a_i *and* α *and* a_j *and* α, *respectively,* $(0 \leq \beta_i, \beta_j \leq \pi)$, $i, j = 1, 2, \ldots, n$.
If further, neither a_i *nor* a_j *is in the same direction as* α, *then for all* θ

$$Cov\left(X_i, X_j \mid \Theta_\alpha = \theta\right) \begin{cases} > 0, \text{if } \beta_{ij} < \pi/2; \\ = 0, \text{if } \beta_{ij} = \pi/2; \\ < 0, \text{if } \beta_{ij} > \pi/2. \end{cases} \tag{3}$$

Moreover, the magnitude of $Cov\left(X_i, X_j \mid \Theta_\alpha = \theta\right)$

is a strictly decreasing function of β_{ij} *for* $\beta_{ij} \in [0, \pi]$ *when* $\left\|a_i\right\| \sin\beta_i$ *and* $\left\|a_j\right\| \sin\beta_j$ *are fixed and* D > 2.

Further, Zhang and Stout (1999) conjecture that moving an item's vector away from the direction in which Θ_α is measured increases the magnitude of its conditional covariance with another fixed item. A demonstration of the results of this theorem most relevant for our purpose can be seen for a different example set of items in Figure 2. Note that in contrast to Figure 1, the representation for the set of items in Figure 2 is in the $\Theta_\alpha^\perp$-space, the latent space orthogonal to the direction in which Θ_α is measured. The four items in Figure 2 are measuring three dimensions, but because they are represented in the $\Theta_\alpha^\perp$-space, only two dimensions are needed for the depicted representation. For now we ignore the results of the theorem related to the sign of the conditional covariance and focus solely on the strictly decreasing association between $Cov(X_i, X_j \mid \Theta_\alpha = \theta)$ and β_{ij}, provided

$$\left\|a_i\right\| \sin\beta_i \text{ and } \left\|a_j\right\| \sin\beta_j$$

are held constant. According to the theorem, increasing the angle between any two items in the $\Theta_\alpha^\perp$-space corresponds to a decrease in their conditional covariance. Thus, in the example in Figure 2, rotating a_1 to the direction of a_1^* results in a decrease in $Cov(X_1, X_3 \mid \Theta_\alpha = \theta)$. Likewise, if it were assumed that the lengths of the vectors representing Items 1, 2, and 3 were all equal, the implied ordering of conditional covariances from smallest to largest would be

$$Cov(X_1, X_2 \mid \Theta_\alpha = \theta) \geq Cov(X_2, X_3 \mid \Theta_\alpha = \theta) \geq Cov(X_1, X_3 \mid \Theta_\alpha = \theta). \tag{4}$$

BEST COPY AVAILABLE

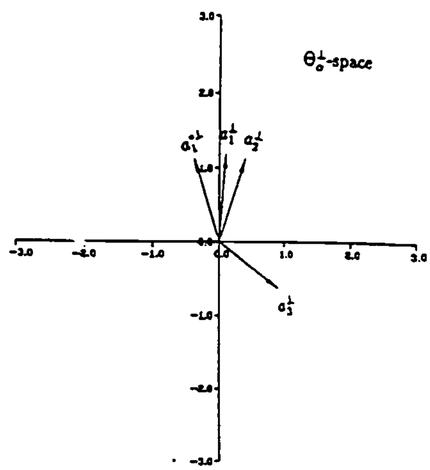


FIGURE 2. *Example of item vector representation in the latent space orthogonal to the direction of best measurement of all the items*

Estimating Conditional Covariances

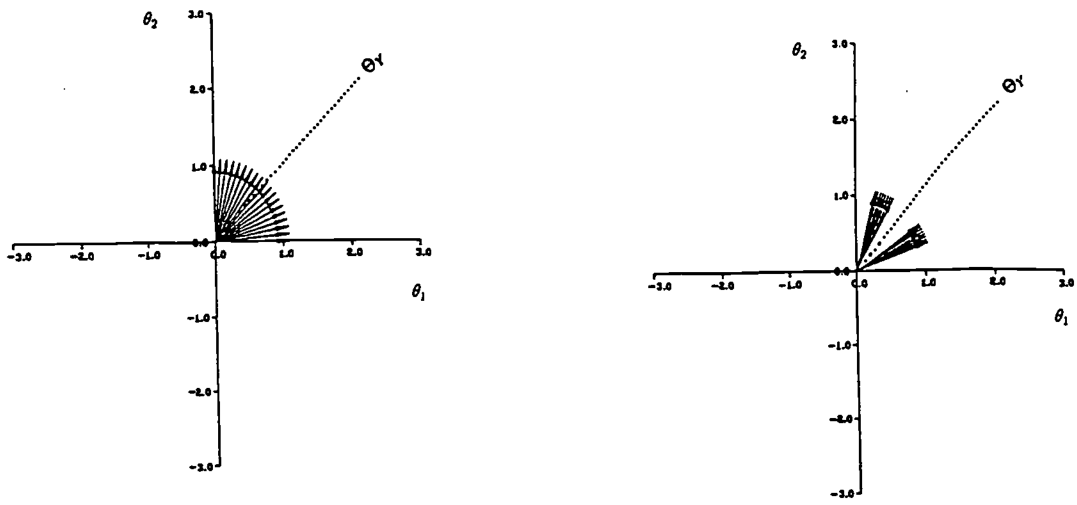
Fundamental to each of the conditional-covariance-based dimensionality tools mentioned earlier is accurate estimation of the conditional covariances, in particular, selection of an appropriate subtest score on which to condition. For both theoretical and practical reasons, the sum score for the remaining items on the test (excluding the two being examined), or

$$S_{ij} = \sum_{l \neq i,\, j} X_l, \tag{5}$$

may be expected to serve as a suitable substitute for θ_Y. Using S_{ij} is advantageous both in being easy and quick to compute and in that, if properly scaled, it approaches θ_Y asymptotically as the number of test items increases. For details on the usual MLE of $Cov(X_i, X_j \mid S_{ij})$ as an estimate of $Cov(X_i, X_j \mid \theta_Y)$ see Kim (1994) and Zhang and Stout (1999).

Measuring the Directional Dissimilarities of Item Pairs

Typically the most informative aspects of item vectors in terms of interpreting test structure are their directions. One example of this is the use of item vector plots to illustrate the amount of simple structure present in a test. This can be assessed by examining the degree to which the items "clump" in a small number of common measurement directions. Figure 3 provides illustrations for two hypothetical tests each of which is two dimensional, but with the second containing a much greater amount of simple structure than the first. We sometimes refer to the structure of Figure 3(b) as approximate simple structure.



(a.) A test displaying complete lack of simple structure. **(b.) A test displaying approximate simple structure.**

FIGURE 3. *Item measurement-direction vectors for two hypothetical two-dimensional tests*

One way in which measurement direction might be inferred is by examining each item's pattern of conditional covariances with all of the other items in the test. For example, given the two-dimensional test displayed in Figure 3(b), it is intuitive that two items whose measurement directions are on the same side of the measurement direction in which test composite θ_Y is measured will have a positive conditional covariance while two items with measurement directions on opposite sides will have a negative conditional covariance. Likewise, in a higher dimensional case, item pairs having positive conditional covariances may be considered to be closer dimensionally than those with negative covariances. (This, in overly simple terms, is what provides a basis for the DETECT procedure.) Therefore, a pair of items from a common dimensional cluster should have very similar patterns of conditional covariances with all of the remaining items in the test. In other words, not only will the item pair itself have a positive conditional covariance, but each item in the pair should also have conditional covariances of the same sign (positive or negative) with all of the remaining items.

In terms of the vector projections in the $\theta_Y^{\perp}$-space (the space orthogonal to the direction in which θ_Y is measured), Zhang's (1996) above-stated theorem suggests that items having common measurement directions should also have common "orderings" of their conditional covariances with the other items. This follows from the fact that items having a common direction also have the same angular distances from the other items. In particular, an item pair measuring the exact same direction in this space should theoretically have the same ordering of conditional covariances with the remaining items from most positive to most negative. Also, if the lengths of the vectors (or their discriminating powers) can be factored out, the actual conditional covariances themselves should be the same (ignoring for now the confounding effects of differences in item difficulty, which will exist in practice).

Given these results, it may be useful to consider an item's association with all other items in attempting to reconstruct its direction. For any particular item i, we define a weighted sum of its estimated conditional covariances with an item j as

$$\widehat{Cov}\left(X_i, X_j\right) = \frac{1}{J}\sum_{k=0}^{n-2} J_k \widehat{Cov}\left(X_i, X_j | S_{ij} = k\right), \tag{6}$$

where J denotes the total number of examinees, J_k is the number of examinees obtaining a raw score $S_{ij} = k$, and n is the number of test items. Then we denote item i's vector of estimated conditional covariances with the remaining items on the test as

$$c_i = \left[\widehat{Cov}(X_i, X_1), \widehat{Cov}(X_i, X_2), \ldots, \widehat{Cov}\left(X_i, X_{(i-1)}\right), \widehat{Cov}(X_i, X_{(i+1)}), \ldots, \widehat{Cov}(X_i, X_n)\right]. \tag{7}$$

Then to compare this vector for a given item i with the corresponding vector for a given item j, we remove from each of the c_i and c_j vectors the element representing its conditional covariance with the other item, and denote the new vector (now of length $n - 2$) $c_{i(j)}$, where the item in parentheses indicates the conditional covariance term removed. As a result, the elements of $c_{i(j)}$ and $c_{j(i)}$ are item i's and j's conditional covariances with all items other than i and j. From these we can construct a commonly used measure of directional similarity between the item pair i, j, namely

$$r_{ij} = \frac{c_{i(j)}^{T} c_{j(i)}}{\left\| c_{i(j)} \right\| \left\| c_{j(i)} \right\|}, \tag{8}$$

which is the product-moment correlation of the elements in the two vectors. Based on Zhang's (1996) theorem, items that measure similar dimensional composites should have large positive, small in magnitude, and large negative conditional covariances with the same sets of items, and, therefore, have a high value for r_{ij}. Items that measure the identical dimensional composite should have an r_{ij} of approximately 1. Items measuring very different dimensional composites will have their strongest positive and negative associations with different sets of items, typically resulting in a negative r_{ij}.

As will be seen next, it will be more useful to consider a measure of dissimilarity between items than one of similarity. Therefore, we compute

$$\delta_{ij} = 1 - r_{ij} \tag{9}$$

as a measure of the degree to which the items measure *different* dimensional composites. This index δ_{ij} assumes values ranging from 0 (when the items measure the same composite) to 2 (when they measure very different composites).

Using MDS to Obtain Orientation of Item Vectors

A useful tool for constructing a representation of psychological stimuli given a proximity matrix such as $\{\delta_{ij}\}$ produced above is multidimensional scaling (MDS). In this case the psychological stimuli being scaled are test items. However, in contrast to the usual MDS approach which attempts to represent dissimilarities between stimuli as distances between points in a plane or hyperplane, our objective is to translate dissimilarities between items into substantively interpretable angular distances between item vectors. Thus, items that have small dissimilarities (implying dimensional closeness) should be close in angular proximity. Two related forms of *constrained* MDS, referred to as circular and spherical scaling (see, Cox & Cox, 1991; Hubert, Arabie, & Meulman 1996), will be used for this purpose. In their usual applications, these methods arrange stimuli on the surface of a circle or sphere (or hypersphere) so that the dissimilarities are represented with respect to distances along the arc of a circle (or surface of a sphere). However, such representations can also in effect be seen as ordering the stimuli so that the differences in their angular directions from the center of the circle or sphere represent dissimilarity. Items arranged so as to have a short arc length between them, correspondingly have a small angle between their vectors (projected from the center), while items located far apart in terms of arc length produce a large angle and, thus, have vectors with very different measurement directions. Thus, circular and spherical scaling techniques provide a way of using MDS to recover angular distances between items.

As is the objective of any multidimensional scaling, we can construct a configuration of points whose interpoint distances d_{ij} represent at some level the degree of dissimilarity δ_{ij} between the objects being scaled. For points constrained to lie on a unit m-dimensional hypersphere, these dissimilarities find their representation in the distances between points along the surface of the hypersphere. The resulting spherical coordinates of a point i in the hyperspherical configuration are given by $(1, \eta_{1i}, \eta_{2i}, \ldots, \eta_{m-1,i})$. In this case, the first element, 1, defines the distance of the point from the origin, while the angle η_{hi} indicates the direction of the vector (from the origin through the point) away from the axis corresponding to dimension $h + 1$. In the two- and three-dimensional cases considered in this paper, these spherical coordinates convert to Cartesian coordinates as $(\sin\eta_{1i}, \cos\eta_{1i})$ and $(\sin\eta_{1i} \sin\eta_{2i}, \cos\eta_{1i} \sin\eta_{2i}, \cos\eta_{2i})$, respectively. Because there is a 1-1 increasing relation between distances along the surface of a sphere (or arc of a circle) and the corresponding Euclidean distances between these points, it is possible to follow the usual MDS procedure of scaling on hyperplanes provided the resulting points are required to lie on a circle or sphere. Euclidean distance between two points on the circle reduces to

$$d_{ij} = \sqrt{2\left(1 - \cos\left(\eta_{1i} - \eta_{1j}\right)\right)}, \tag{10}$$

while in the spherical case

$$d_{ij} = \left(2 - 2\sin\eta_{2i}\sin\eta_{2j}\cos\left(\eta_{1i} - \eta_{1j}\right) - 2\cos\eta_{2i}\cos\eta_{2j}\right)^{1/2}. \tag{11}$$

For a specific MDS configuration, a frequently used measure of how well the resulting distances between points of the configuration, d_{ij} represent the given dissimilarities, δ_{ij}, is given by a measure of *stress*, S (Kruskal, 1964a, 1964b), computed as

$$S = \left(\frac{\sum_{(i,\,j)} \left(d_{ij} - \hat{d}_{ij}\right)^2}{\sum_{(i,\,j)} d_{ij}^2} \right)^{\frac{1}{2}} \tag{12}$$

where

$$\hat{d}_{ij} = \hat{f}\left(\delta_{ij}\right).$$

Here $\hat{f}$ provides the "primary" monotone least squares regression of d_{ij} on δ_{ij}. That is, we minimize

$$\sum \left(d_{ij} - \hat{d}_{ij}\right)^2$$

over all choices of monotone $\hat{f} \geq 0$ while requiring for all $\delta_{ij} < \delta_{kl}$ that $d_{ij} \leq d_{kl}$. A "best" configuration is one that has minimum stress, with $S = 0$ indicating perfect fit. This type of scaling solution is considered *nonmetric* because the function $\hat{f}$ can take any monotonic form and, thus, the resulting configuration displays the dissimilarities at an ordinal level only.

An approach to producing a configuration that minimizes the stress value S using the method of steepest descent is discussed by Cox and Cox (1991) and implemented in the program MDSCAL-T (Cox & Cox, 1996). One practical complication with the procedure is that results are often influenced by the presence of local minima, especially when using poor initial starting configurations. Empirical results suggest that starting with an initial MDS configuration that first sorts the items into clusters (say based on a cluster analysis) and then positions the clusters an equal distance from each other on the surface of the sphere may be more effective in leading to an optimal solution.

An alternative approach that avoids this problem when scaling on a circle is provided by Hubert et al. (1996). Their method gives emphasis to finding first an optimal ordering of the stimuli on a circle, and then spacing the stimuli so as to accurately reproduce the dissimilarities using a least squares optimization strategy. The Hubert et al. approach also differs from the Cox and Cox approach in that it provides a metric solution, namely one in which the reproduced distances are related linearly to the original dissimilarities, and, thus, is slightly more restrictive.

Assessing and Representing the Adequacy of the Representation

It has been traditional in multidimensional scaling to focus on stress S as an index of fit in the nonmetric case and *variance accounted for* (VAF) in the metric case. Nevertheless, it is perfectly reasonable to apply both indices in the metric and nonmetric cases. VAF is equal to the proportion of variance in the dissimilarities accounted for by distances in the final configuration. In the metric case, this is equivalent to the square of the correlation between the corresponding dissimilarities and distances, or $r^2(\delta, d)$. In the nonmetric case, we compute it as $r^2(\hat{d}, d)$, assuming that an appropriate rescaling of dissimilarities is that defined by the function $\hat{f}$ (recall that $\hat{d}_{ij} = \hat{f}\left(\delta_{ij}\right)$). Likewise, in the metric case, a stress value may be computed as

$$S = \left(\frac{\sum_{(i,\,j)} \left(d_{ij} - \hat{d}_{ij}\right)^2}{\sum_{(i,\,j)} d_{ij}^2} \right)^{\frac{1}{2}}, \tag{13}$$

where $\hat{f}$ is now restricted as the linear transformation that minimizes this expression, and $1 \leq i < j \leq n$.

Practically speaking, either of these indices (stress or VAF) permits one to decide if an extra dimension (i.e., moving from a circle to a sphere, or sphere to hypersphere) would be useful in improving the representation.

It may also frequently be the case that the resulting representation reproduces a particular item's dissimilarities with the remaining items more effectively than the representation does for some other item. In particular, items that measure in the same direction that θ_Y is measured will not have a meaningful representation in the final configuration. Along these lines, our decision to control for the varying lengths of the "true" vector projections (indicating their discrimination strength with respect to common secondary dimensions) by standardizing the c_i vectors to construct δ_{ij} likewise removes from the representation the degree to which items measure in the same direction as θ_Y. We will, however, want to build some notion of the length of these vectors back into our final representation.

Analogous to the multidimensional logistic model approach, we represent the degree to which items measure common dimensions orthogonal to the direction in which θ_Y is measured through the length of the item vector. The accuracy of the representation for an individual item may be determined using VAF for each item separately, or $\text{VAF}_i = r^2(\delta_{ij}, d_{ij})$, across all $1 \le j \le n$, $j \ne i$. Further, the varying levels of discriminating power for each item may be most evident in the magnitudes of their conditional covariances with the other items. Therefore, a value such as the ratio of the mean absolute conditional covariance between item i and the rest of the items over the mean absolute conditional covariance among all of the items, can provide information as to the relative discriminating strength of each item in the $\theta_Y^{\perp}$ plane. To allow both of these elements to influence vector length, we make vector length proportional to the product of VAF_i and this ratio. It is important to note that our notion of discrimination as captured by vector length in this representation is, therefore, different from that defined earlier. In particular, in addition to the usual discrimination of the items, it is also dependent both on the relative angular distance of the item from θ_Y, as well as where the other items lie in the representation (i.e., whether it lies in a "clump" of items measuring a common secondary dimension).

Analysis of Simulated Data

To examine whether the scaling procedure is useful in recovering multidimensional test structure, data were simulated for several hypothetical multidimensional tests using Reckase's multidimensional extension of the two-parameter logistic model. Unless otherwise specified, for each of the cases below, 3,000 examinees were simulated from a multivariate normal ability distribution

$$\left(\mu_1 = \mu_2 = \ \dots \ = \mu_k = 0, \sigma_1 = \sigma_2 = \dots = \sigma_k = 1\right) \tag{14}$$

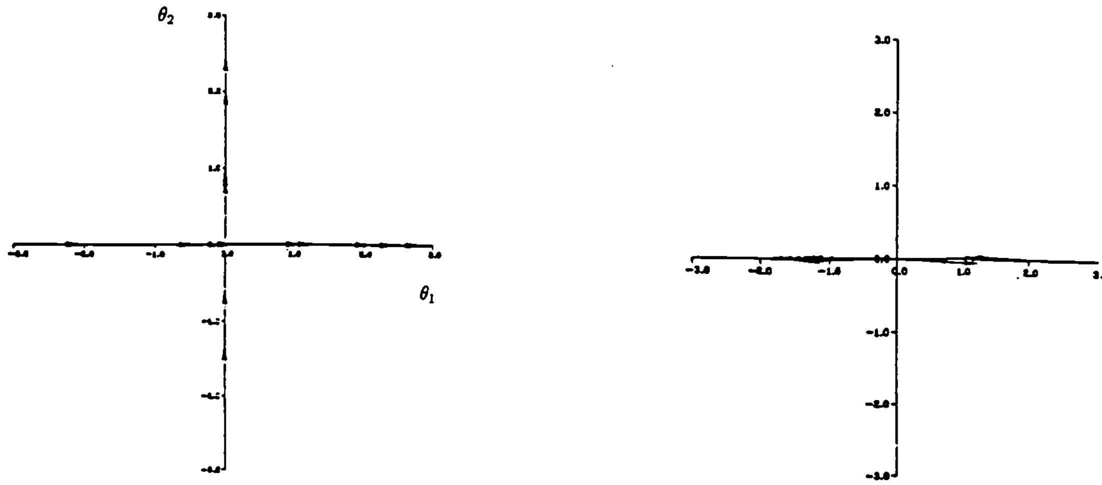
with a correlation of .5 between each pair of k dimensions. In each of the two 2-dimensional cases and in each of the four 3-dimensional cases, two-part figures are provided to compare the simulated test structure with the estimated test structure from the MDS results. The test structure used to simulate data is represented in θ-space in accordance with Reckase's multidimensional logistic representation, and the estimated test structure is represented by the resulting scaling configuration, which corresponds to $\theta_Y^{\perp}$-space. Both difficulty d_i and discrimination were varied for all items with $d_i \sim N(0, .75)$ and

$$\sqrt{\sum_{k=1}^{D} a_{ik}^2} \sim N(1, .3), \tag{15}$$

where D is the number of dimensions.

Two-Dimensional Case

In the two-dimensional case, a $\theta_Y^{\perp}$-space representation (which is unidimensional because $dim(\theta_Y^{\perp}) = dim(model) - 1$) should theoretically position all of the items in only one of two directions, depending on which of the two dimensions the item more predominantly measures. The first example considered is the simple structure case in which all items are "pure" measures of either Dimension 1 or Dimension 2. Figure 4(a) displays a case in which an equal number of items was used to measure each dimension. Also apparent is the variability in difficulty and discrimination among items measuring each dimension. Using the circular scaling technique (Hubert et al., 1996) combined with our approach for reproducing vector lengths, the reproduced structure is shown in Figure 4(b). The VAF value of .995 indicates a good recovery of the computed angular dissimilarity among items. Recovery of the actual angular differences is further supported by the fact that each of the sets of items measuring a common dimension in Figure 4(a) is oriented in a common direction in Figure 4(b). Thus, the scaling technique effectively sorted the items into their appropriate clusters.



(a.) Simulated dimensionality structure in θ-space. **(b.) Estimated dimensionality structure in $\theta_Y^{\perp}$ space.**

FIGURE 4. *Comparison of simulated and estimated dimensionality structures. Two-dimensional simple structure. Equal number of items per dimension*

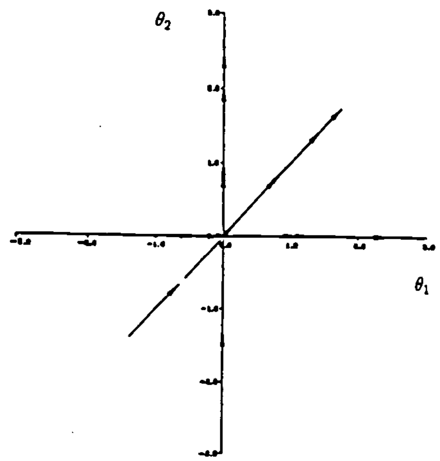
In some cases, it may be that there exist items that are in themselves pure measures of θ_Y. Figure 5(a) illustrates a structure similar to that in Figure 4(a), but with a cluster of six items that measure

$$\frac{\theta_1 + \theta_2}{2}, \tag{16}$$

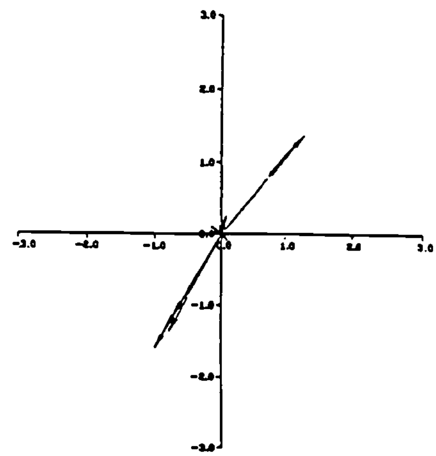
which is approximately in the same direction in which θ_Y is measured. As a result there exist a total of three clusters of items, although the test itself is still two-dimensional. Because performance on the items measuring

$$\frac{\theta_1 + \theta_2}{2}$$

is predominantly determined only by θ_Y (i.e., not by $\theta_Y^{\perp}$), their estimated conditional covariances with the other items will be very close to 0 and the projections of their vectors on the $\theta_Y^{\perp}$ plane will be small. Thus, there should not be any meaningful association between these items and the others in the $\theta_Y^{\perp}$-space. Figure 5(b) provides the results. Again items that were pure measures of either of the two dimensions are correctly sorted so as to point in different directions. The VAF in this representation of .844 is slightly lower than in the previous case, due to the "noise" contributed by items measuring only θ_Y. Each of these pure θ_Y items is represented by a vector having almost zero length. As will always be the case in this type of representation, the vectors suggest which dimensions other than θ_Y are measured by the individual test items.



(a.) Simulated dimensionality structure in θ-space.



(b.) Estimated dimensionality structure in $\theta_Y^\perp$ space.

FIGURE 5. *Comparison of simulated and estimated dimensionality structures. Two-dimensional simple structure. Items partitioned equally between measuring* θ_1, θ_2, *and* $\frac{\theta_1+\theta_2}{2}$

It should be mentioned that the presence or absence of a simple structure is not detectable in the $\theta_Y^\perp$-space for the two-dimensional case. Instead, vector length captures the varying degree to which an item measures a dimension other than θ_Y. As a result, in the strictly two-dimensional case, it should always be the case (approximately at least) that each of the vectors of non-zero length points in just one of two directions. In this sense, our use of circular scaling in the two-dimensional case actually places the representation in a higher dimensional space than is necessary. Equivalent representations to those in Figures 4(b) and 5(b) can be obtained using a unidimensional scaling representation.

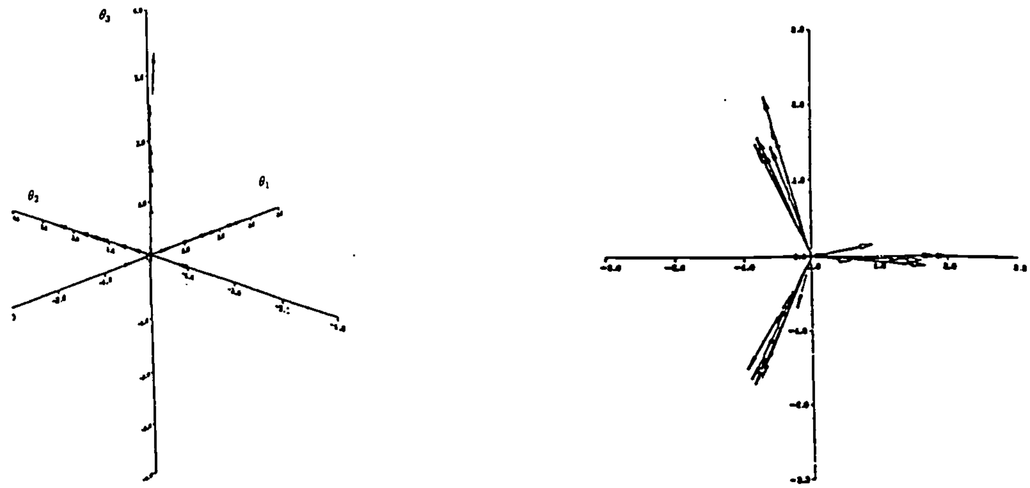
Three-Dimensional Case

The three-dimensional case permits a more thorough investigation of the flexibility of our scaling methodology. Specifically, the capacity to examine conditions of simple structure versus lack of simple structure and to examine varying levels of association between dimensionally distinct item clusters requires at least three dimensions.

Figure 6(a) illustrates the three-dimensional simple structure case analogous to the two-dimensional structure in Figure 4(a). For clarity, the items in each cluster are represented by the same type of arrowhead (closed & shaded, closed & unshaded, open). This three-dimensional representation of the multidimensional logistic model is a direct extension of the two-dimensional case, with the length of the vector now proportional to

$$\sqrt{a_1^2+a_2^2+a_3^2} \tag{17}$$

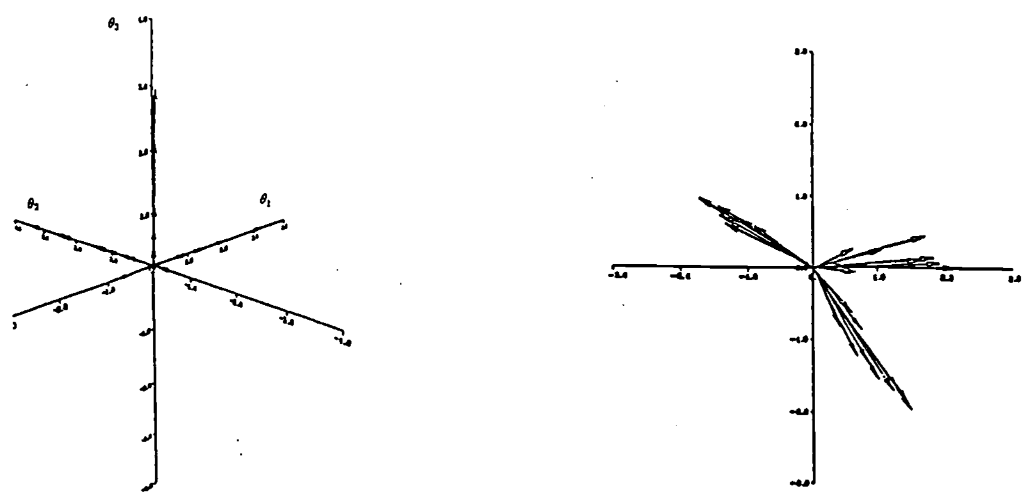
and the orientation of the vector again defined by the direction of maximum information. In this case, different arrowheads define items from the different dimensional clusters. Figure 6(b) shows the result based on applying the scaling procedure. Again the VAF is quite high (.993). The length of the vectors, although associated with the discriminating power of the item (r = .43), are also heavily influenced by the item difficulty. This is not surprising given the known effect of differences in difficulty on the conditional covariance. Nevertheless, difficulty appears to have little effect on the vector directions at least when there is not a strong association between difficulty and dimensionality.



(a.) Simulated dimensionality structure in θ-space. (b.) Estimated dimensionality structure in $\theta_V^{\perp}$ space.

FIGURE 6. *Comparison of simulated and estimated dimensionality structures. Three-dimensional simple structure. Equal number of items per dimension. Equal correlations between the dimensions*

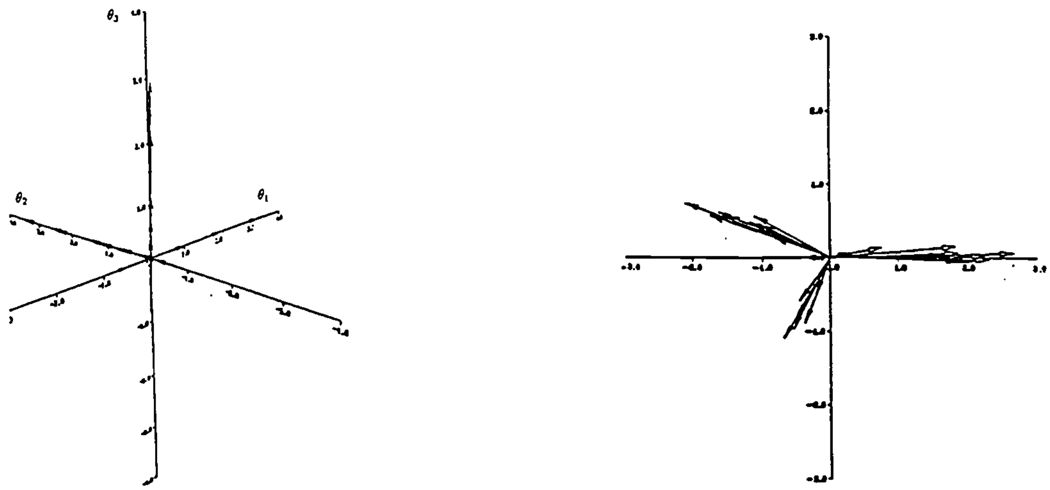
Figures 7, 8, and 9 display alternative three-dimensional structures. In Figure 7(a), the correlation between dimensions is varied so that Dimensions 1 and 2 are correlated .7, while the other pairs remain correlated .5. Note that the resulting configuration in Figure 7(b) (VAF = .978) correctly orients the two more highly associated clusters closer together, while keeping their relative distances from the third cluster equal. Thus, it appears that an interpretation of the similarity of clusters may be possible using the scaling approach.



(a.) Simulated dimensionality structure in θ-space. (b.) Estimated dimensionality structure in $\theta_V^{\perp}$ space.

FIGURE 7. *Comparison of simulated and estimated dimensionality structures. Three-dimensional simple structure. Equal number of items per dimension. Correlation of 0.7 between θ_1 and θ_2 while the other correlations are 0.5*

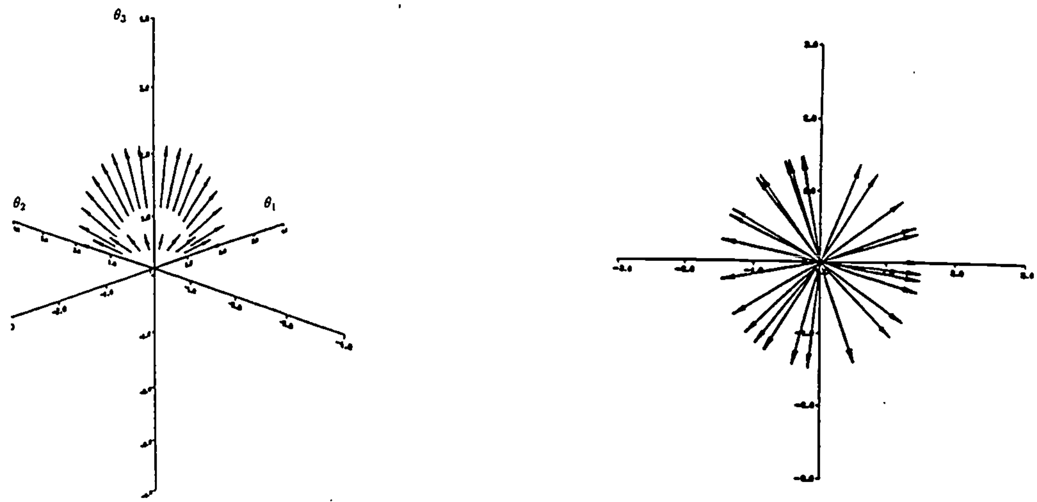
In Figure 8(a), the number of items in each cluster has been varied so that there are 5, 10, and 15 items measuring each of the respective dimensions. This appears to have only a small effect on the resulting configuration in Figure 8(b) (VAF = .958). Again all items are correctly sorted into their respective clusters, but this time the clusters themselves are oriented so as to indicate a slightly larger difference between the two larger-sized clusters.



(a.) Simulated dimensionality structure in θ-space. **(b.) Estimated dimensionality structure in $\theta_V^{\perp}$ space.**

FIGURE 8. *Comparison of simulated and estimated dimensionality structures. Three-dimensional simple structure. Five items measuring θ_1, 10 measuring θ_2, and 15 measuring θ_3. Equal correlations between the dimensions*

In Figure 9(a), a three-dimensional test totally lacking simple structure is represented. In this case there are a total of 30 items, with each item measuring a composite of two dimensions. For each pair of dimensions there are 10 items spread equally in 10-degree increments, producing a fan-like representation (similar to that in Figure 3[a]) within each two-dimensional plane. Although item discrimination and difficulty were varied in the simulation, vector lengths are drawn here as equal to simplify the representation. As desired, in Figure 9(b) the resulting configuration (VAF = .950) emphasizes the lack of simple structure and orients each set of vectors measuring a different two-dimensional composite in a different 120-degree region. Thus, unlike DETECT, our procedure seems to produce interpretable results in cases where there is a complete lack of simple structure.



(a.) Simulated dimensionality structure in θ-space.

(b.) Estimated dimensionality structure in $\theta_V^{\perp}$ space.

FIGURE 9. *Comparison of simulated and estimated dimensionality structures. Three dimensions with complete lack of simple structure. All dimensions equally well measured. Equal correlations between the dimensions*

Four-Dimensional Case

For tests that are four-dimensional, a more appropriate representation will likely be obtained in a three-dimensional space using the spherical scaling approach. Table 1 displays the four-dimensional discrimination parameter vectors for a test comprised of 48 items. For this test, some degree of simple structure is present, although there are a variety of different dimensional composites being measured. The scaling solution provided in Table 2 corresponds to a stress value of .025, and appears to effectively represent the original structure. Note that in this solution spherical coordinates are defined with respect to two angles, the first, denoted η_1, being the orientation of the vector with respect to the X-axis in the XY plane (measured in degrees) and the second, denoted η_2, being the orientation of the vector with respect to the Z-axis (or vertical axis). Items having a common dimensional structure are oriented in common directions. This can be seen, for example, in the subsets of items that are pure measures of each of the four dimensions (item groups {1–6}, {13–18}, {25–30}, {37–48}), where each subset points in a different direction, and all items within each subset point in a common direction. Further, items that are composites of these dimensions typically have directions located in between the directions of these respective subsets. For example, Item 10, which is largely a composite of Dimensions three and four, is directed to a location equidistant from item cluster {1–6} and item cluster {25–30}, which measure, respectively, Dimensions three and four.

TABLE 1
Item discrimination and difficulty level used for the simulated four-dimensional data

Item	a_1	a_2	a_3	a_4	d
1	0.00	0.00	0.00	1.00	-1.24
2	0.00	0.00	0.00	1.00	0.35
3	0.00	0.00	0.00	1.00	-0.26
4	0.00	0.00	0.00	1.00	-1.67
5	0.00	0.00	0.00	1.00	-0.23
6	0.00	0.00	0.00	1.00	1.74
7	0.00	0.00	0.50	0.87	1.55
8	0.00	0.00	0.50	0.87	-0.99
9	0.00	0.00	0.50	0.87	-0.67
10	0.00	0.38	0.66	0.64	1.75
11	0.27	0.27	0.66	0.64	-0.53
12	0.38	0.00	0.66	0.64	0.45
13	0.00	1.00	0.00	0.00	-0.05
14	0.00	1.00	0.00	0.00	0.44
15	0.00	1.00	0.00	0.00	1.28
16	0.00	1.00	0.00	0.00	-2.13
17	0.00	1.00	0.00	0.00	-1.22
18	0.00	1.00	0.00	0.00	-0.46
19	0.00	0.87	0.00	0.50	-1.34
20	0.00	0.87	0.00	0.50	0.63
21	0.00	0.87	0.00	0.50	0.81
22	0.00	0.64	0.38	0.66	0.62
23	0.27	0.64	0.27	0.66	-0.05
24	0.38	0.64	0.00	0.66	-1.44
25	0.00	0.00	1.00	0.00	-1.59
26	0.00	0.00	1.00	0.00	1.47
27	0.00	0.00	1.00	0.00	-0.37
28	0.00	0.00	1.00	0.00	0.39
29	0.00	0.00	1.00	0.00	-2.13
30	0.00	0.00	1.00	0.00	0.95
31	0.00	0.50	0.87	0.00	-0.46
32	0.00	0.50	0.87	0.00	-2.07
33	0.00	0.50	0.87	0.00	1.33
34	0.00	0.66	0.64	0.38	-1.03
35	0.27	0.66	0.64	0.27	0.51
36	0.38	0.66	0.64	0.00	-0.47
37	1.00	0.00	0.00	0.00	-0.45
38	1.00	0.00	0.00	0.00	1.06
39	1.00	0.00	0.00	0.00	0.34
40	1.00	0.00	0.00	0.00	1.23
41	1.00	0.00	0.00	0.00	0.35
42	1.00	0.00	0.00	0.00	1.52
43	1.00	0.00	0.00	0.00	-0.44
44	1.00	0.00	0.00	0.00	0.38
45	1.00	0.00	0.00	0.00	-0.25
46	1.00	0.00	0.00	0.00	-0.94
47	1.00	0.00	0.00	0.00	1.73
48	1.00	0.00	0.00	0.00	1.11

TABLE 2
Estimated item measurement directions in $\theta_Y^{\perp}$-space from the MDS solution for the four-dimensional simulated data

Item	η_1	η_2
1	-55	142
2	-53	122
3	-37	134
4	-50	123
5	-51	136
6	-41	124
7	33	118
8	27	121
9	35	131
10	27	101
11	53	113
12	87	106
13	-80	81
14	-75	80
15	-79	76
16	-77	84
17	-70	77
18	-82	75
19	-69	90
20	-72	91
21	-75	92
22	-46	102
23	-64	104
24	-84	99
25	75	84
26	77	85
27	73	85
28	74	84
29	73	82
30	76	84
31	65	79
32	68	77
33	63	82
34	16	85
35	53	70
36	81	51
37	6	-71
38	13	-90
39	12	-80
40	9	-83
41	12	-79
42	6	-81
43	12	-82
44	11	-76
45	5	-76
46	2	-83
47	9	-79
48	10	-83

With nonmetric MDS, it is typically useful to refer to a Shepard plot (Shepard, 1962), which both represents the function $\hat{f}$, as well as a plot of the d_{ij}s versus the δ_{ij}s. Figure 10 displays this plot for the simulated data and solution shown in Table 2. Note the closeness of the plotted points to the line representing $\hat{f}$, indicating a good representation of the original dissimilarities in the solution. Also, the nature of $\hat{f}$ suggests that the angular distances in the final representation can be interpreted as almost linear with respect to the dissimilarities.

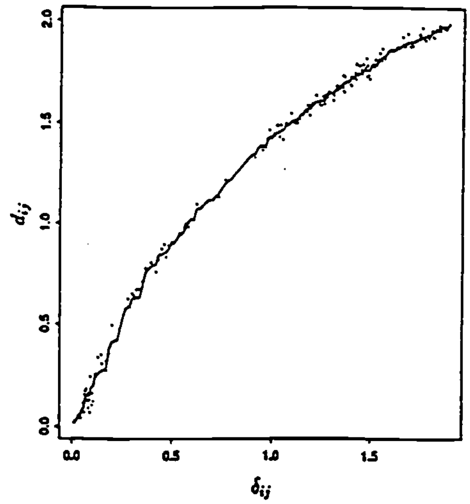


FIGURE 10. *Shepard diagram corresponding to the MDS solution for the four-dimensional simulated data*

Analysis of Real Data

Previous analyses of the Law School Admission Test (LSAT) suggest that the forms can be considered to be predominantly measured by two correlated dimensions, one measured by items in an analytical reasoning (AR) section and the other measured collectively by reading comprehension (RC) and logical reasoning (LR) section items (Camilli, Wang, & Fesq, 1995; De Champlain, 1995, 1996; Douglas, Kim, Roussos, Stout & Zhang, 1999; Stout, et al. 1996; Wilson & Powers, 1994). Separate dimensionality analyses of the AR and RC sections further suggest that these two sections each have passage-based structures; that is, items based on a common reading passage tend to be more dimensionally similar than those based on different passages (Camilli, Wang, & Fesq, 1995; Douglas, et al. 1999; Stout, et al. 1996) In an attempt to replicate these findings, separate analyses were performed for each of the AR and RC sections. Because each of the sections is based on 4 passages, it was anticipated that a three-dimensional representation using the spherical scaling approach would be most appropriate, recalling that we expect, in general, to need $D - 1$ dimensions in our representation if D equals the number of dominant dimensions.

Results for the AR section (the same data examined by Stout et al.) using the spherical scaling procedure are shown in Table 3. For purposes of clarity, only the angles for each item are presented rather than the actual vector representation. In this case, the items from the four item sets separate very cleanly into four clusters (Items 1–7, 8–13, 14–19, and 20–24), equally spaced in terms of angular distance. In fact, the directions for items from common clusters are nearly identical in all cases. What makes this possible is the very unrestrictive nature of the function $\hat{f}$ linking the δ_{ij} s to the d_{ij} s. These results are in complete agreement with the HCA/CCPROX and DETECT analyses reported in Stout et al.

TABLE 3

Estimated item measurement directions in $\theta_Y^{\perp}$-space from the MDS solution for the analytical reasoning test data

Item	η_1	η_2
1	131	0
2	122	0
3	119	0
4	124	0
5	119	0
6	120	0
7	129	0
8	147	111
9	147	111
10	147	111
11	147	111
12	147	111
13	147	111
14	27	110
15	27	110
16	27	110
17	27	110
18	27	110
19	27	110
20	268	110
21	268	110
22	268	110
23	268	110
24	268	110

Perhaps the most interesting results occurred in the analysis of the RC section. This test also consists of items based on 4 passages, and so was also analyzed using the spherical approach, with results shown in Table 4. Previous analyses of this RC section using HCA/CCPROX and DETECT had yielded conflicting results (Stout et al.). The four-cluster solution from the HCA/CCPROX analysis divided the section into four clusters with each corresponding to exactly one of the four item sets that went with the four passages. The DETECT analysis, on the other hand, suggested a three-cluster solution. Two of the DETECT clusters corresponded to two separate item sets, but the third cluster was the union of the items corresponding to the two remaining item sets. As shown in Table 4, the current results indicate a four-cluster solution in agreement with the previous HCA/CCPROX analysis. Furthermore, the new information yielded by the current analysis offers a compelling explanation for the difference in the two previous analyses results. The current analysis indicates that two of the clusters, Items 1 to 7 and Items 8 to 15, though dimensionally distinct in terms of measurement direction, are much closer together than any other pair of clusters. The near proximity of those two clusters to each other made it difficult for DETECT to tell the difference between them. Unlike cluster analysis and MDS, which rely only on ordinal information, the DETECT analysis relies on accurate measurement of the signs of the conditional covariances. Because the conditional covariance estimator used in these techniques may be statistically biased, DETECT may sometimes have difficulty distinguishing clusters that are very close together. Because HCA and MDS rely on only ordinal information, any statistical bias in the estimators of the conditional covariances does not adversely affect these procedures and they remain effective in distinguishing such clusters. The substantive reason as to why these two clusters were close together can be readily found in that the two passages corresponding to these clusters are both "science-based." This finding further lends support to the capability of the circular/spherical scaling approach to represent varying degrees of associations between item clusters.

TABLE 4
Estimated item measurement directions in $\theta_Y^{\perp}$-space from the MDS solution for the reading comprehension test data

Item	η_1	η_2
1	145	66
2	157	73
3	140	74
4	153	71
5	153	64
6	151	79
7	144	74
8	216	74
9	213	80
10	210	72
11	215	82
12	214	79
13	211	83
14	210	79
15	219	83
16	331	5
17	346	7
18	310	20
19	35	17
20	346	12
21	6	130
22	0	131
23	6	124
24	0	130
25	359	125
26	9	124
27	8	132
28	8	133

It is perhaps best to interpret the AR and RC results with respect to the estimated function $\hat{f}$ associating the dissimilarities to the distances. Figures 11 and 12 provide Shepard plots for the AR and RC analyses, respectively. Note that in the AR representation there are only two different potential distances, so that for every item from a common cluster, although varying in their actual dissimilarities, the distance is 0, while every item pair from different clusters has a distance of approximately 1.75. Because in this case item pairs from common clusters had δ_{ij}s below the minimum δ_{ij} for item pairs from different clusters, the fit is also perfect, resulting in a VAF $\approx$ 1.0 and stress $\approx$ 0.00. Again, it should be recalled that the representations obtained preserve the ordinal associations between items with respect to the δ_{ij}s. Therefore, in this case we limit our interpretation to the observation that items from common passages are more highly associated than items from different passages. We cannot assess from this Shepard diagram the relative associations between the passage-based clusters.

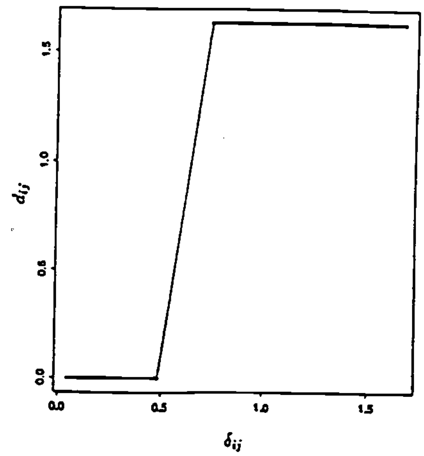


FIGURE 11. *Shepard diagram corresponding to the MDS solution for the analytical reasoning test data*

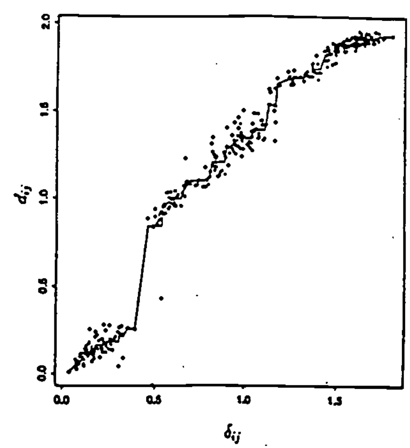


FIGURE 12. *Shepard diagram corresponding to the MDS solution for the reading comprehension test data*

In the reading comprehension case, there is some variability of points about the $\hat{f}$ line suggesting that fit is not perfect, but still quite adequate (VAF = .962, stress = .032). Also, in contrast to the AR case, there are a variety of different distances represented. Note in particular the group of item pair points having distances of approximately 1.0. These represent the items from the two clusters corresponding to the two science-based reading passages. These two clusters had a closer association than the other clusters.

Conclusion

The results of this preliminary study suggest that inter-item conditional covariances may provide a useful basis for constructing a meaningful item-level test geometry. Compared to other proximity measures developed for the purpose of dimensionality assessment (see Roussos, Stout, & Marden, 1998), the dissimilarity index used in this paper is unique in its use of patterns of conditional covariances with the other items in the test. The purpose of this is to control for the length of vector projections in the $\theta_Y^{\perp}$-space and build an index that is only sensitive to dissimilarity in vector direction. Future study will be needed to provide a more careful investigation of the properties of this dissimilarity index and its relationship to the measures proposed by Roussos, et al. (1998). More specifically, the use of the index proposed here in the context of HCA (either by itself or in conjunction with DIMTEST) should be investigated.

More investigation is also needed into the nature of the association between the index and the true angle between items in the $\theta_Y^{\perp}$-space. Because other factors besides this angle may contribute to the dissimilarity index, it is important to gain a more thorough interpretation of the reproduced angles (i.e., the item geometry) in light of the index used. Further, a more precise understanding of the effects of difficulty on conditional covariances and how to best control for it in constructing the representation would be warranted.

As used in this paper, circular and spherical MDS procedures seemingly provide within MDS a step in the direction of providing a factor-analytic type of representation of test structure. It may be possible within this context to make more meaningful comparisons between these two classes of dimensionality-reduction techniques, which have historically been considered quite disparate. One advantage to the use of MDS is the wide variety of possible dissimilarity indices with which it can be used. It may be useful to examine the performance of this scaling method with others that are also sensitive to the dimensional structure of the items.

Finally it will be important to link this approach with the other nonparametric dimensionality tools mentioned earlier. To the extent that all are conditional covariance based, it might be expected that they perform similarly in their descriptions of test structure. As demonstrated in this paper, one advantageous feature of the scaling approach relative to HCA and DETECT is its capacity to represent varying levels of association among clusters. Moreover, these other tools can also serve as an aid in constructing a scaling-based representation. For example, HCA/CCPROX and DETECT can help define the dimensionality of the space in which a representation is most appropriate.

References

Camilli, G., Wang, M. M., & Fesq, J. (1995). The effects of dimensionality on equating the Law School Admission Test. *Journal of Educational Measurement, 32*, 79–96.

Cox, T. F., & Cox, M. A. A. (1991). Multidimensional scaling on a sphere. *Communications in Statistics- Theory and Methods, 20*, 2943-2953.

Cox, T. F., & Cox, M. A. A. (1996). *Multidimensional scaling*. New York: Marcel-Dekker.

De Champlain, A. F. (1995). *Assessing the effect of multidimensionality on LSAT equating for subgroups of test takers* (Statistical Report 95-01). Newtown, PA: Law School Admission Council.

De Champlain, A. F. (1996). The effect of multidimensionality of IRT true-score equating for subgroups of examinees. *Journal of Educational Measurement, 33*, 181–201.

Douglas, J., Kim, H. R., Roussos, L., Stout, W., & Zhang, J. (1999). *LSAT dimensionality analysis for the December 1991, June 1992, and October 1992 Administrations* (Statistical Report 95-05). Newtown, PA: Law School Admission Council.

Hubert, L., Arabie, P., & Meulman, J. (1996). *Linear and circular unidimensional scaling for symmetric proximity matrices*. Research Report 96-02, Department of Data Theory, Leiden University.

Kim, H. R. (1994). *New techniques for the dimensionality assessment of standardized test data*. Unpublished doctoral dissertation. University of Illinois, Champaign-Urbana.

Knol, D. L., & Berger, P. F. (1991). Empirical comparison between factor analysis and multidimensional item response models. *Multivariate Behavioral Research, 26*, 457–477.

Kruskal, J. B. (1964a). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika, 29*, 1–27.

Kruskal, J. B. (1964b). Nonmetric multidimensional scaling: a numerical method. *Psychometrika, 29*, 115–129.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems.* Hillside, NJ: Lawrence Erlbaum.

McDonald, R. P. (1967). Nonlinear factor analysis. *Psychometric Monographics* (No. 15).

McDonald, R. P. (1985). Unidimensional and multidimensional models for item response theory. In D.J. Weiss (Ed), *Proceedings of the 1982 Item Response Theory and Computer Adaptive Testing Conference.* Minneapolis, University of Minnesota.

Pashley, P. J. (1995). *LSAT computerized assessment research agenda.* Newtown, PA: Law School Admission Council.

Pashley, P. J. (1997). *Computerized LSAT research agenda: Spring 1997 update.* Newtown, PA: Law School Admission Council.

Reckase, M. D. (1985). The difficulty of test items that measure more than one ability. *Applied Psychological Measurement, 9,* 401-412.

Reckase, M. D. (1997). A linear logistic multidimensional model. In W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp.271-296). New York: Springer-Verlag.

Reckase, M. D., & McKinley, R. L. (1991). The discriminating power of items that measure more than one dimension. *Applied Psychological Measurement, 14,* 361–373.

Roussos, L. A. (1992). *Hierarchical agglomerative clustering program users manual.* Unpublished manuscript. Department of Statistics, University of Illinois at Urbana-Champaign.

Roussos, L. A., Stout, W. F., & Marden, J. (1998). Using new proximities with hierarchical cluster analysis to detect multidimensionality. *Journal of Educational Measurement, 35,* 1-30.

Shepard, R. N. (1962). The analysis of proximities: Multidimensional scaling with an unknown distance function. *Psychometrika, 27,* 125–140, 219–246.

Stout, W. F. (1987). A nonparametric approach for assessing latent trait dimensionality. *Psychometrika, 52,* 589–617.

Stout, W. F. (1990). A new item response theory modeling approach with applications to unidimensional assessment and ability estimation. *Psychometrika, 55,* 293–326.

Stout, W. F., Habing, B., Douglas, J., Kim, H. R., Roussos, L. A., & Zhang, J. (1996). Conditional covariance based nonparametric multidimensionality assessment. *Applied Psychological Measurement, 20,* 331–354.

Wang, M. (1988, April). *Measurement bias in the application of a unidimensional model to multidimensional item-response data.* Paper presented at the annual meeting of the American Education Research Association, New Orleans.

Wilson, K. M., & Powers, D. E. (1994). *Factors in performance on the Law School Admission Test* (Statistical Report 93-04). Newtown, PA: Law School Admission Council.

Zhang, J. (1996). *Some fundamental issues in item response theory with applications.* Unpublished doctoral dissertation. University of Illinois at Urbana-Champaign.

Zhang, J., & Stout, W. F. (1999). Conditional covariance structure of generalized compensatory multidimensional items, with applications. *Psychometrika, 64,* 129–152.

NOTICE

Reproduction Basis

X	This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.
	This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").